Inductive biases for deep learning of higher-level cognition

Year: 2020 | Citations: 411 | Authors: Anirudh Goyal, Yoshua Bengio

Abstract

A fascinating hypothesis is that human and animal intelligence could be explained by a few principles (rather than an encyclopaedic list of heuristics). If that hypothesis was correct, we could more easily both understand our own intelligence and build intelligent machines. Just like in physics, the principles themselves would not be sufficient to predict the behaviour of complex systems like brains, and substantial computation might be needed to simulate human-like intelligence. This hypothesis would suggest that studying the kind of inductive biases that humans and animals exploit could help both clarify these principles and provide inspiration for AI research and neuroscience theories. Deep learning already exploits several key inductive biases, and this work considers a larger list, focusing on those which concern mostly higher-level and sequential conscious processing. The objective of clarifying these particular principles is that they could potentially help us build AI systems benefiting from humans' abilities in terms of flexible out-of-distribution and systematic generalization, which is currently an area where a large gap exists between state-of-the-art machine learning and human intelligence.